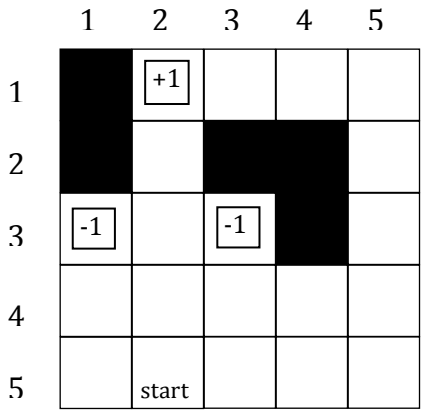
Markov Decision Processes

Consider the following environment. Use value iteration to compute the utility of each state. The states with squares inside are terminal states.

The reward for each terminal state is given in the boxes. The reward for any non-terminal state is R(s) = -0.01.

The agent can perform the following actions: UP, DOWN, LEFT, RIGHT. The agent's actuators are stochastic. There is an 80% chance that the action will execute correctly and a 10% chance that the agent will navigate 90-degrees clockwise and 10% chance that the agent will navigate 90-degrees counter-clockwise.

The utility values for terminal states are the same as the reward values. The initial utility values for all non-terminal states, s, are $U_0(s) = 0.0$. Use a discount factor of 0.5.



Problem: Compute $U_1(s)$ and $U_2(s)$. Use batch value iteration.

Solution:

States are listed as (y, x).

In the case of a tie for the policy, the solution arbitrarily breaks ties and only lists one.

$U_1(s)$:	$Pi_1(s)$:
(1, 2): 1 (1, 3): 0.39 (1, 4): -0.01 (1, 5): -0.01, (2, 2): 0.39, (2, 5): -0.01, (3, 1): -1, (3, 2): -0.11, (3, 3): -1, (3, 5): -0.01, (4, 1): -0.01, (4, 2): -0.01, (4, 3): -0.01, (4, 4): -0.01, (4, 5): -0.01, (5, 1): -0.01, (5, 2): -0.01, (5, 3): -0.01, (5, 4): -0.01, (5, 5): -0.01,	(1, 2): None (1, 3): left (1, 4): ['up', 'down', 'left', 'right'] (1, 5): ['up', 'down', 'left', 'right'] (2, 2): up (2, 5): ['up', 'down', 'left', 'right'] (3, 1): None (3, 2): ['up', 'down'] (3, 3): None (3, 5): ['up', 'down', 'left', 'right'] (4, 1): down (4, 2): ['up', 'down', 'left', 'right'] (4, 3): down (4, 4): ['up', 'down', 'left', 'right'] (4, 5): ['up', 'down', 'left', 'right'] (5, 1): ['up', 'down', 'left', 'right'] (5, 2): ['up', 'down', 'left', 'right'] (5, 3): ['up', 'down', 'left', 'right'] (5, 4): ['up', 'down', 'left', 'right'] (5, 5): ['up', 'down', 'left', 'right']

$U_2(s)$:	$Pi_2(s)$:
(1, 2): 1, (1, 3): 0.429 (1, 4): 0.145 (1, 5): -0.015 (2, 2): 0.429 (2, 5): -0.015 (3, 1): -1 (3, 2): 0.046 (3, 3): -1 (3, 5): -0.015 (4, 1): -0.015 (4, 2): -0.015 (4, 3): -0.015 (4, 4): -0.015 (4, 5): -0.015 (5, 1): -0.015 (5, 2): -0.015 (5, 3): -0.015 (5, 4): -0.015 (5, 5): -0.015	(1, 2): None (1, 3): left (1, 4): left (1, 5): ['up', 'down', 'left', 'right'] (2, 2): up (2, 5): ['up', 'down', 'left', 'right'] (3, 1): None (3, 2): up (3, 3): None (3, 5): ['up', 'down', 'left', 'right'] (4, 1): down (4, 2): down (4, 3): down (4, 4): ['up', 'down', 'left', 'right'] (4, 5): ['up', 'down', 'left', 'right'] (5, 1): ['up', 'down', 'left', 'right'] (5, 2): ['up', 'down', 'left', 'right'] (5, 3): ['up', 'down', 'left', 'right'] (5, 4): ['up', 'down', 'left', 'right'] (5, 5): ['up', 'down', 'left', 'right']